Consistent Estimation of Pricing Kernels from Noisy Price Data

Vladislav Kargin*

February 1, 2008

Abstract

If pricing kernels are assumed non-negative then the inverse problem of finding the pricing kernel is well-posed. The constrained least squares method provides a consistent estimate of the pricing kernel. When the data are limited, a new method is suggested: relaxed maximization of the relative entropy. This estimator is also consistent.

Keywords: ε-entropy, non-parametric estimation, pricing kernel, inverse problems.

JEL: G12,G13, C14

MSC (2000) 62G08, 45Q05, 65J22

*Cornerstone Research, 599 Lexington Avenue, New York, NY 10022, USA; slava@bu.edu

1 Introduction

Modern finance theory postulates that the price of a security is an integral of its future payoff multiplied by a pricing kernel:

$$S(x, \theta) = \int F(x', \theta) p(x, x') dx'. \tag{1}$$

Here S represents the security price, x and x' current and future values of stochastic factors relevant for pricing the security, θ a non-stochastic parameter, F the future payoff, and p the pricing kernel. The pricing kernel is of great interest to finance theory because it sheds light on investors' preferences over current and delayed consumption. Practitioners are also interested in the pricing kernel because it helps in pricing new securities, finding mispriced assets, and managing risk[1]. Non-surprisingly, when it was discovered that the pricing kernel can be recovered from option prices[2], financial economists en masse went agog inventing new and better methods for estimating the pricing kernel.[3] Many of the methods, however, are heuristic and lack a rigorous proof of consistency. This paper focuses on providing a simple proof of consistency for the constrained least squares and a modified maximum entropy methods.

Mathematically, the pricing kernel estimation is an inverse problem. A linear operator maps a set of functions ("pricing kernels") into another set of functions ("security prices"), and the problem is to invert this operator. Often the problem is additionally complicated by the fact that prices are observed only for a discrete set of securities and contaminated with noise. This kind of inverse problems frequently appears in diverse areas of applied mathematics and thoroughly studied [4].

The pricing kernel estimation is, however, special and what makes it special is that the pricing kernel must be non-negative to prevent the existence of systematic arbitrage opportunities.

This restriction on the operator's domain helps a lot. Without it, the inverse problem is ill-posed, that is, the pricing operator does not have a continuous inverse. Intuitively, small changes in security prices could lead to large changes in the estimate of the pricing kernel. In addition, without this restriction, the least squares method of estimation is inconsistent. The pricing kernel selected by the least squares would fit the prices exactly but would not converge to the true pricing kernel. In contrast, non-negativity of the pricing kernel makes the corresponding inverse problem well-posed and the least squares method consistent.

[1] See, for example, applications in Jackwerth (2000), Ait-Sahalia and Lo (2000), and Rosenberg and Engle (2002).

[2] By Breeden and Litzenberger (1978) and Banz and Miller (1978), and revived by Rubinstein (1994).

[3] An incomplete list includes Jackwerth and Rubinstein (1996), Avellaneda et al. (1997), Sőderlind and Swensson (1997), Melick and Thomas (1997), Ait-Sahalia and Lo (1998), Avellaneda (1998), Jackwerth (2000), Ait-Sahalia and Duarte (2003), and Bondarenko (2003).

[4] See reviews in Tikhonov and Arsenin (1977), O'Sullivan (1986), and Engl (2000).

The key to the well-posedness is that for pricing purposes it is enough to estimate the distribution function of the pricing kernel: cumulative pricing kernel. These functions form a Banach space with respect to the uniform convergence topology, and we can apply one of the Banach theorems: A continuous one-to-one operator on a Banach space has a continuous inverse. Consequently, the inverse problem is well-posed.

What can be said about consistency? By well-posedness, estimating the kernel can be reduced to estimating price function from noisy observations: the map from price functions to pricing kernels is continuous and cannot inflate the error of estimation. Luckily, the problem of estimating the price function is the classic problem of non-parametric estimation of a regression function, and for this problem the conditions of the least squares consistency are well known. It turns out that they are satisfied provided the pricing kernel is non-negative. Intuitively, additional information about the structure of pricing kernels prevents overfitting of the regression function and forces consistent convergence of the estimates. Together with well-posedness, this implies that the constrained least squares estimates the cumulative pricing kernel consistently.

While asymptotically consistent, the constrained least squares may, however, perform unsatisfactorily in small samples. It is because this method ignores prior information about the pricing kernel. One way to abate the problem is to include in the objective function a term that measures distance from the prior information pricing kernel. This idea leads to a method that combines advantages of both the least squares and the maximal entropy methods. The method maximizes the weighted sum of relative entropy and the mean squared pricing error. With suitably chosen parameters, this method is also consistent.

Let me briefly describe the related literature. The maximum entropy method for estimating pricing kernel was developed by Buchen and Kelly (1996) and Stutzer (1996), following a suggestion in Rubinstein (1994), and elaborated by Avellaneda et al. (1997), Avellaneda (1998), and Fritelli (2000). These papers typically assume that the securities are priced correctly but only a scarce discrete set of prices is known. For the alternative case of large amount of noisy data, methods of pricing kernel estimation based on smoothing or other ideas were developed by Jackwerth and Rubinstein (1996), Ait-Sahalia and Lo (1998), Jackwerth (2000), and Bondarenko (2003) among others. Implicitly, these papers address the problem of ill-posedness of kernel estimation by the classic method of regularization. This paper is different because it shows that on the restricted domain of non-negative kernels the problem is well-posed and so does not need additional regularization.

Ait-Sahalia and Duarte (2003) estimate the pricing kernels by smoothing the *constrained* least squares estimator, and refer to the statistical literature for the proof of consistency. In this paper, we provide an explicit proof of the constrained least squares consistency and consider another modification of the method based on the idea of entropy distance minimization.

The rest of the paper is organized as follows. Section 2 reminds the basics of the theory of pricing kernels. Section 3 shows that the problem of finding the non-decreasing cumulative pricing kernel is well-posed. Section 4 demonstrates

consistency of the least squares. Section 5 explains how the idea of the maximum entropy can be used to improve the least squares method, and proves the consistency of the modification. Section 6 concludes.

2 What is the pricing kernel?

The pricing kernel, p, is a function of stochastic factors that allows pricing securities by using their future payoff functions:

$$S(x,\theta) = \int F(x',\theta)p(x,x')dx'. \tag{2}$$

Often, the choice of units in which the stochastic factors are measured is arbitrary, so we can normalize the initial level of factors: $x = 1$. Slightly abusing notation, we will denote $p(1,x)$ as $p(x)$. Let us define *cumulative pricing kernel* as follows:

$$P(x) = \int_{-\infty}^{x} p(t)dt. \tag{3}$$

With these notations, the pricing formula can be rewritten in a more convenient form:

$$S(\theta) = \int F(x,\theta)dP(x). \tag{4}$$

Non-negativity of the pricing kernel, implied by the absence of arbitrage opportunities (Harrison and Kreps (1979)), translates into monotonicity of the cumulative pricing kernel: $P(x)$ is non-decreasing. In addition, the price of the security that have a unit payoff is finite, so $P(x)$ is bounded.

We will be interested in pricing kernels that depend only on one factor. For example, when the class of securities consists of options written on another security, this factor is the price of the underlying security. We can further simplify the problem by noting that payoff of most derivative securities that occur in practice can be represented as a linear combination of underlying security and security that has a non-zero payoff only if the price of underlying is less than a certain bound, B. Therefore, we can concentrate on pricing the derivatives with finite support, and then by integration by parts we have the following pricing formula:

$$S(\theta) = -\int_{0}^{B} P(x)dF(x,\theta), \tag{5}$$

where B is such that $F(x,\theta) = 0$ for $x > B$.

The next lemma shows that estimating cumulative pricing kernel is sufficient for pricing purposes. Let $P_n(x)$ be an estimate of $P(x)$. Let S_n be the corresponding price of the derivative from (5).

Lemma 1 *Suppose $F(x)$ has bounded variation and P_n converges to P in uniform metric as n goes to ∞. Then S_n converges to S.*

Proof:

$$\begin{aligned} |S_n - S| &= \left| \int_0^B \left[P_n(x) - P(x)\right] dF(x) \right| \leq \int_0^B \left|P_n(x) - P(x)\right| dF(x) && (6) \\ &\leq C \left\| P_n(x) - P(x) \right\|_\infty , && (7) \end{aligned}$$

where C is the total variation of $F(x)$. QED.

Consider now how we can estimate the pricing kernel. Typically, it is done by using the prices of puts. A European put with strike K is a security that will pay:

$$F(x, K) = \max(K - x, 0) \tag{8}$$

at the expiration date if the price of the underlying security is x on that date. Then, the price of the put with strike K is

$$\begin{aligned} S(K) &= \int_0^K (K - x)\, dP(x) && (9) \\ &= \int_0^K P(x) dx. && (10) \end{aligned}$$

The operator of interest is then

$$A : P(x) \to S(K) = \int_0^K P(x) dx. \tag{11}$$

In a more general setting, we are interested in the inverse problem defined by operator

$$A_F : P(x) \to S(\theta) = -\int_0^B P(x) dF(x, \theta). \tag{12}$$

As Lemma 1 shows, operator A_F is continuous in the uniform metric (L^∞). We are interested in knowing whether its inverse is continuous, that is, if small deviations in prices can lead to large deviations in pricing kernel. We also need to know if the pricing kernel can be consistently estimated from noisy and discrete data. These problems are handled in the next two sections.

3 Pricing problem is well-posed.

A problem $Ax = y$ is called well-posed if the operator A has a continuous inverse. It is implicit in this definition that the operator is given with its domain, and that topologies in both the range and the domain are specified: the same operator may be ill-posed on one domain and well-posed on another one. The concept of well-posedness originated in mathematical physics by Hadamard as a tool to select the linear problems that could arise from a physical problem. Later, however, it was discovered that many important problems are ill-posed and the methods of their solution were derived (Tikhonov and Arsenin (1977), O'Sullivan (1986), Engl (2000)).

If no restrictions on pricing kernels were imposed, then the operator in (11) would correspond to an ill-posed problem. Indeed, it is easy to see ill-posedness from the following example:

$$\alpha \cos(\beta x) \to \int_0^K \alpha \cos(\beta x) dx = \frac{\alpha}{\beta} \sin(\beta K). \tag{13}$$

Consider the uniform convergence metric on both the domain and the range of the operator. If we set $\beta = \alpha/\varepsilon$, then the norm of the function on the left-hand side is constant: $\|\alpha \cos(\beta x)\|_\infty = \alpha$, but its image can be made arbitrarily close to zero $\|F(\alpha \cos(\beta x))\|_\infty = \|\varepsilon \sin(\beta K)\|_\infty \leq \epsilon$: therefore the inversion operator acts discontinuously.

However, for the restricted domain of cumulative pricing kernels we have the following theorem:

Theorem 1 *If A_F is injective then it defines a well-posed problem on the space of all non-decreasing continuous functions with uniform convergence topology.*

Proof: In uniform convergence topology the space of non-decreasing continuous functions is complete. If A_F is injective, then it defines a continuous one-to-one correspondence between this space and its image. The conclusion of the theorem follows because of one of the Banach theorems (see for example Theorem 11 in Chapter 15 of Lax (2002)): A linear operator that establishes a continuous one-to-one correspondence between two complete normed linear spaces has a continuous inverse. QED.

Corollary 1 *Operator A from (11) defines a well-posed problem on the space of all non-decreasing continuous functions with uniform convergence topology.*

Proof: Since A is injective, Theorem 1 can be applied. QED.

In practice securities prices are known up to an error. This error includes bid-ask spread, non-stationarity in the pricing kernel, market inefficiencies and so on. We will use Theorem 1 and Corollary 1 as tools to prove that as the amount of data grows the constrained least squares estimates the pricing kernel consistently.

4 Estimation by least squares is consistent.

Let $\{\Omega, \Sigma, \Pr\}$ be a probability space and ε_i be a sequence of independent identically distributed random variables with zero expectation and finite variance. Let also x_i be a sequence of points located between 0 and B, which has a positive density on $[0, B]$. We will say that the constrained least squares estimates function f from set $\mathcal{F}$ consistently in norm $\|\cdot\|$ relative to operator A if for any δ, with probability 1 there exists such N_0 that for $N \geq N_0$, there exists

$$f_N = \arg\max_{\widehat{f} \in \mathcal{F}} \sum_{i=1}^{N} \left(Af(x_i) - A\widehat{f}(x_i) \right)^2, \tag{14}$$

and $\|f - f_N\| < \delta$. In other words, with probability 1, the sequence of estimates f_N converges to the true function.

Here we are interested in the set, $\mathcal{P}$, of non-decreasing, continuous, bounded functions on interval $[0, B]$. We use the uniform convergence topology and operator A from (11).

Theorem 2 *The constrained least squares estimates any function in $\mathcal{P}$ consistently in L^∞ relative to operator A.*

Proof: Let $\mathcal{R}$ be the image of $\mathcal{P}$ under operator A. Then $\mathcal{R}$ is the set of convex, increasing, continuous, bounded functions. Because of Theorem 1, operator A has a continuous inverse from $\mathcal{R}$ to $\mathcal{P}$. Consequently, consistency of estimating functions from $\mathcal{P}$ relative to operator A is equivalent to consistency of estimating unmodified functions from $\mathcal{R}$. For $\mathcal{R}$, we can apply classic results for the non-parametric estimation of convex function. In particular, according to the main Theorem in Hanson and Pledger (1976), convex functions can be estimated consistently in L^∞ norm by the constrained least squares method. QED.

For a more general operator A_F from (12) we have a similar theorem, which, however, needs a more advanced technique and comes to a weaker conclusion.

Let us call payoff function $F(x, \theta)$ *uniformly Lipschitz* in θ if

$$|F(x, \theta_1) - F(x, \theta_2)| \leq C\,|\theta_1 - \theta_2|\,, \tag{15}$$

where C does not depend on x. Also let us call $F(x, \theta)$ *uniformly bounded in variation*, if its total variation over $x \in [0, B]$ is bounded by a constant that does not depend on θ.

Theorem 3 *If $F(x, \theta)$ is uniformly Lipschitz in θ and uniformly bounded in variation, and A_F is injective, then the constrained least squares estimates any function in $\mathcal{P}$ consistently in L^2 relative to operator A_F.*

In the proof we will again aim to prove that any function in $\mathcal{R} = A_F(\mathcal{P})$ can be estimated consistently by the constrained least squares. We are going to do it by referring to a theorem in van de Geer (1987). First, let us introduce several additional concepts. Let X be a set of functions on $\mathbb{R}^k$ and let $M_n(\delta, \mathcal{X})$ be the minimal number of elements in a $\delta-$covering of set $\mathcal{X}$, if the distance is measured by the norm

$$\|f\|_n = \frac{1}{n}\sum_{i=1}^{n}\left[f(x_i)\right]^2. \tag{16}$$

Then $\delta-$*entropy* of a set is defined as

$$N_n(\delta, \mathcal{X}) = \frac{1}{n}\log M_n(\delta, \mathcal{X}). \tag{17}$$

Note that $\delta-$*entropy* depends on the choice of points x_i. We assume that they are distributed randomly according to a measure, μ, that has a positive continuous density on $[0, B]$. Then let us call a set of functions *entropically thin* if for

any δ

$$N_n(\delta, \mathcal{X}) \rightarrow_\mu 0 \text{ as } n \rightarrow \infty, \tag{18}$$

where convergence is in probability. Intuitively, a set of functions is entropically thin if all its functions can be well approximated by functions from a relatively "small" subset.[5]

Next, a class of functions, $\mathcal{X}$, is called *uniformly square integrable* if

$$\lim_{C\rightarrow\infty} \sup_{f\in\mathcal{X}} \int_{|f|>C} f^2 dx = 0. \tag{19}$$

A somewhat weaker version of van de Geer's result is sufficient for our purposes. It says that if a set $\mathcal{X}$ is uniformly square integrable and entropically thin, then the constrained least squares method is L^2-consistent.

Proof of Theorem 3: Any $S \in \mathcal{R}$ is representable as

$$S(\theta) = -\int_0^B P(x) dF(x, \theta). \tag{20}$$

Since set $\mathcal{P}$ is uniformly bounded and $F(x,\theta)$ is uniformly bounded in variation, set $\mathcal{R}$ is also uniformly bounded. Consequently, it is uniformly square integrable.

Similarly, since $F(x,\theta)$ is uniformly Lipschitz in θ, and $\mathcal{P}$ is uniformly bounded, set $\mathcal{R}$ is uniformly Lipschitz:

$$\begin{aligned} |S(\theta_1) - S(\theta_2)| &= \left| \int_0^B \left[F(x,\theta_1) - F(x,\theta_2) \right] dP(x) \right| && (21) \\ &\leq \int_0^B C_1 \left| \theta_1 - \theta_2 \right| dP(x) && (22) \\ &\leq BC_1C_2 \left| \theta_1 - \theta_2 \right|. && (23) \end{aligned}$$

Consequently, by Lemma 3.3.1 in van de Geer (1987) $\mathcal{R}$ is entropically thin. Therefore, van de Geer's theorem can be applied and the constrained least squares estimator is L^2-consistent.

QED.

The conditions of Theorem 3 are not very restrictive. For example, the set of payoff functions for puts is uniformly Lipschitz:

$$\left| \max\{K_1 - x, 0\} - \max\{K_2 - x, 0\} \right| \leq |K_1 - K_2|. \tag{24}$$

It is also clearly uniformly bounded in variation over $x \in [0, B]$, provided that we consider only a bounded set of the strikes: $\max\{K - x, 0\} \leq \overline{K} \equiv \max K$. Finally, the pricing operator, A, is injective if $\overline{K} \geq B$. Therefore, Theorem 3 is applicable.

[5]The concept of entropy in relation to totally bounded sets of functions was introduced by Kolmogorov and Tikhomirov (1959). It was applied to the problem of consistency in non-parametric estimation by Vapnik and Červonenkis (1981). For a textbook presentation, see Pollard (1984).

While asymptotically consistent, the constrained least squares may perform poorly in small samples. It fails to take into account such possible prior beliefs as that the pricing kernel is smooth, or unimodal, or that it is approximately proportional to an infinitely divisible probability distribution, etc. In the next section we consider a modification of the method of constrained least squares that allows to take into account the prior information.

5 Relaxed Maximum Relative Entropy Method

In this section we will for simplicity restrict the discussion to the case when the pricing kernel is estimated from the put prices. Relaxed maximum entropy method penalizes both the degree to which the model fails in explaining the price data and the model's deviation from a prior model:

$$\widehat{P}(x) = \arg\min_{P(x)} \left\{ \frac{1}{N} \sum_{i=1}^{N} (S(K_i) - S_i)^2 + \lambda_N \int_0^B \ln \frac{dP(x)}{dP_0(x)} dP(x) \right\}, \quad (25)$$

where S_i is the observed price of the put with strike K_i,

$$S(K) = AP(x) \equiv \int_0^K P(x)dx, \quad (26)$$

and $P_0(x)$ is a prior cumulative pricing kernel.

Recall that the regular maximum entropy method is described by the following minimization problem:

$$\widehat{P}_{ME}(x) = \arg\min_{P(x)} \left\{ \int_0^B \ln \frac{dP(x)}{dP_0(x)} dP(x) \text{ s.t. } S(K_i) = S_i \text{ for each } i \right\}. \quad (27)$$

If prices are contaminated with noise, then the regular maximum entropy may run into difficulties with the existence of the solution and is unlikely to be consistent. In the relaxed maximum entropy method, constraints are not rigid, they are substituted with a penalizing term in the objective function. Consequently, the solution is guaranteed to exist. What about consistency?

Theorem 4 *There is such a sequence of positive constants λ_N, that the relaxed maximum entropy method estimates the cumulative pricing kernel, $P(x)$, consistently in L^2 norm.*

Proof: By a lemma below, there is such a sequence $\lambda_N \to 0$ that as $N \to \infty$, the solution of the problem

$$\min_{S(K)} \left\{ \frac{1}{N} \sum_{i=1}^{N} (S(K_i) - S_i)^2 + \lambda_N \int_0^B \ln \frac{dP(x)}{dP_0(x)} dP(x) \right\} \quad (28)$$

with probability 1 approaches in L^2 norm the solution of the constrained least squares problem:

$$\min_{S(K)} \frac{1}{N} \sum_{i=1}^{N} (S(K_i) - S_i)^2 \text{ s.t. } \partial_K^2 S \geq 0. \tag{29}$$

by Theorem 2, as $N \to \infty$, the solution of the constrained least squares problem with probability 1 approaches the true pricing function $S(K)$.

By the standard diagonal process argument there is such a sequence of λ_N, that the solution of

$$\min_{S(K)} \left\{ \frac{1}{N} \sum_{i=1}^{N} (S(K_i) - S_i)^2 + \lambda_N \int_0^B \ln \frac{dP(x)}{dP_0(x)} dP(x) \right\} \tag{30}$$

approaches in L^2 the true function $S(K)$ as $N \to \infty$. Because the differentiation is a continuous operator on the set of convex non-decreasing functions, $P(x)$ is also estimated consistently in L^2. QED.

In the proof of Theorem 4, we have used the following Lemma. Consider the problem:

$$\min_{R \in \mathcal{R}} \{F_N(R) + \lambda_N G(R)\},$$

where F_N and G are continuous functionals of $R(x)$. Let R_N be solution of the problem, and $\widehat{R}_N$ be the solution for $\lambda = 0$. Let the sequence of functionals F_N be called *proper* on $\mathcal{R}$ if for any ε we can find such δ that for all sufficiently large N, and $R \in \mathcal{R}$, condition $F_N(R) - F_N(\widehat{R}_N) < \delta$ implies that $\left\| R - \widehat{R}_N \right\|_{L^2} \leq \varepsilon$.

Lemma 2 *If $\{F_N\}$ is proper on $\mathcal{R}$, then there exists such a sequence λ_N that $R_N - \widehat{R}_N$ converges to zero.*

Proof: Take an ε and select δ and N_0 as in the definition of properness; then for any $R \in \mathcal{R}$ and any $N \geq N_0$ from $\left\| R - \widehat{R}_N \right\|_{L^2} > \varepsilon$ it follows that $F_N(R) - F_N(\widehat{R}_N) \geq \delta$. On the other hand, from continuity of F_N it follows that we can find such ε_1 that $\left\| R - \widehat{R}_N \right\|_{L^2} < \varepsilon_1$ implies $F_N(R) - F_N(\widehat{R}_N) \leq \delta/2$. Also, since G is continuous, we can find such an R' inside the ε_1-neighborhood of $\widehat{R}_N$ that $|G(R') - G(\widehat{R}_N)| < c$. Consequently we can find such λ that $\lambda G(R') < \delta/2$. Then it is clear that the maximizer of $F_N(R) + \lambda G(R)$ cannot be outside of the $\varepsilon-$neighborhood of $\widehat{R}_N : R'$ would improve on it.

Thus, for any ε, there is N_0 and λ such that for $N \geq N_0$ the solution of $\min\{F_N + \lambda G\}$ is in $\varepsilon-$neighborhood of $\widehat{R}_N$. QED.

This Lemma can be used the proof of Theorem 4 because the functional

$$F_N(R) \equiv \frac{1}{N} \sum_{i=1}^{N} (R(K_i) - S_i) \tag{31}$$

is proper. Indeed, note that this functional has a nice special property:

Lemma 3 *If* $F_N(R) - F_N(\widehat{R}_N) < \varepsilon$ *then* $F_N(R - \widehat{R}_N) < \varepsilon$.

Proof: Let $R = \widehat{R}_N + \delta R$. Since F_N is a quadratic form, we can define a corresponding bilinear product:

$$(f, g) = \frac{1}{2}\left\{F_N(f+g) - F_N(f) - F_N(g)\right\}, \tag{32}$$

We claim that

$$(\widehat{R}_N, \delta R) \geq 0. \tag{33}$$

Indeed, since the set of convex non-decreasing functions, $\mathcal{R}$, is convex, $R_\alpha \equiv \widehat{R}_N + \alpha\delta R \in \mathcal{R}$ for any $\alpha \in [0,1]$. Consequently, if (33) were violated we could find such α that $(R_\alpha, R_\alpha) < (\widehat{R}_N, \widehat{R}_N)$, which would contradict optimality of $\widehat{R}_N$. Using (33), we can write:

$$F_N(\delta R) + F_N(\widehat{R}_N) \leq F_N(R) < F_N(\widehat{R}_N) + \varepsilon, \tag{34}$$

and $F_N(\delta R) < \varepsilon$. QED.

So, to obtain properness of $\{F_N\}$ it remains to prove that from $F_N(R - \widehat{R}_N) < \varepsilon$ for all large N we can conclude that $\left\|R - \widehat{R}_N\right\|_{L_2} \leq \varepsilon$. By assumption, points $\{x_i\}$ are distributed with density $\rho(x) \geq k > 0$ on $[0, B]$. Then we can use the following lemma:

Lemma 4 *If* $f(x)$ *is non-negative and has finite variation on* $[0, B]$ *then from*

$$\frac{1}{N}\sum_{i=1}^{N} f(x_i) \leq \varepsilon \text{ for any } N \geq N_0, \tag{35}$$

it follows that

$$\int_0^B f(x)dx \leq \frac{\varepsilon}{k}. \tag{36}$$

Proof: The sum converges to $\int_0^B f(x)\rho(x)dx$. Therefore,

$$\int_0^B f(x)dx = \int_0^B \frac{f(x)}{\rho(x)}\rho(x)dx \leq \frac{1}{k}\int_0^B f(x)\rho(x)dx \leq \frac{\varepsilon}{k}. \tag{37}$$

QED.

Properness of functional $F_N(x)$ follows from this lemma applied to $f(x) = \left(R(x) - \widehat{R}_N(x)\right)^2$.

6 Conclusion

It is proved that the mapping from the set of non-decreasing cumulative pricing kernels to security prices corresponds to a well-posed inverse problem, and that

the constrained least squares method provides a consistent estimator of the cumulative pricing kernel.

It is also suggested that in small samples the performance of the constrained least squares can be improved by a modification that takes into account that the pricing kernel should be close to a certain prior kernel. It is proved that this method is consistent.

References

Ait-Sahalia, Y. and J. Duarte (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics 116*, 9–47.

Ait-Sahalia, Y. and A. Lo (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance 53*, 499–547.

Ait-Sahalia, Y. and A. Lo (2000). Nonparametric risk management and implied risk aversion. *Journal of Econometrics 94*, 9–51.

Avellaneda, M. (1998). Minimum-relative-entropy calibration of asset-pricing models. *International Journal of Theoretical and Applied Finance 1*(4), 447–472.

Avellaneda, M., C. Friedman, R. Holmes, and D. Samperi (1997). Calibrating volatility surfaces via relative-entropy minimization. *Applied Mathematical Finance 4*(1), 37–64.

Banz, R. and M. Miller (1978). Prices of state-contingent claims: some estimates and applications. *Journal of Business 51*(4), 653–672.

Bondarenko, O. (2003). Estimation of risk-neutral densities using positive convolution approximation. *Journal of Econometrics 116*, 85–112.

Breeden, D. T. and R. H. Litzenberger (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business 51*(4), 621–651.

Buchen, P. W. and M. Kelly (1996). The maximum entropy distribution of asset inferred from option prices. *Journal of Financial and Quantitative Analysis 31*(1), 143–159.

Engl, H. W. (2000). Inverse problems and their regularization. In *Computational mathematics driven by industrial problems (Martina Franca, 1999)*, Volume 1739 of *Lecture Notes in Math.*, pp. 127–150. Berlin: Springer.

Fritelli, M. (2000). The minimal entropy martingale measure and the valuation problem in incomplete markets. *Mathematical Finance 10*(1), 39–52.

Hanson, D. L. and G. Pledger (1976). Consistency in concave regression. *Annals of Statistics 4*(6), 1038–1050.

Harrison, M. J. and D. M. Kreps (1979, June). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory 2*(3), 381–408.

Jackwerth, J. (2000). Recovering risk aversion from option prices and realized returns. *Review of Financial Studies 13*, 433–451.

Jackwerth, J. and M. Rubinstein (1996). Recovering probability distributions from option prices. *Journal of Finance 51*, 1611–1631.

Kolmogorov, A. N. and V. M. Tikhomirov (1959). ϵ-entropy and ϵ-capacity of sets in function spaces. *Uspekhi Matematicheskih Nauk 14*, 3–86. English translation in Amer. Math. Soc. Transl. 17, 277-364, 1961.

Lax, P. D. (2002). *Functional Analysis*. John Wiley and Sons, Inc.

Melick, W. R. and C. P. Thomas (1997). Recovering an asset's implied pdf from option prices: an application to crude oil during the gulf crisis. *Journal of Financial and Quantitative Analysis 32*(1), 91–115.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistical Science 1*(4), 502–518.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer-Verlag, New York.

Rosenberg, J. and R. Engle (2002). Empirical pricing kernels. *Journal of Financial Economics 64*, 341–372.

Rubinstein, M. (1994). Implied binomial trees. *Journal of Finance 49*(3), 771–817.

Sőderlind, P. and L. Swensson (1997). New techniques to extract market expectations from financial instruments. *Journal of Monetary Economics 40*(2), 383–429.

Stutzer, M. (1996). A simple nonparametric approach to derivative security valuation. *Journal of Finance 51*(5), 1633–1652.

Tikhonov, A. N. and V. Y. Arsenin (1977). *Solutions of ill-posed problems*. John Wiley and Sons. With an introduction by Fritz John.

van de Geer, S. (1987). A new approach to least-squares estimation, with applications. *Annals of Statistics 15*(2), 587–602.

Vapnik, V. N. and A. Y. Červonenkis (1981). Necessary and sufficient condition for the uniform convergence of means to their expectations. *Theory of Probability and Its Applications 26*, 532–553.